(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) **Patentschrift**

(10) **DE 196 36 739 C1**

(51) Int. Cl.[6]: G 10 L 5/06

(21) Aktenzeichen: 196 36 739.5-53
(22) Anmeldetag: 10. 9. 96
(43) Offenlegungstag: —
(45) Veröffentlichungstag der Patenterteilung: 3. 7. 97

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

(73) Patentinhaber:
Siemens AG, 80333 München, DE

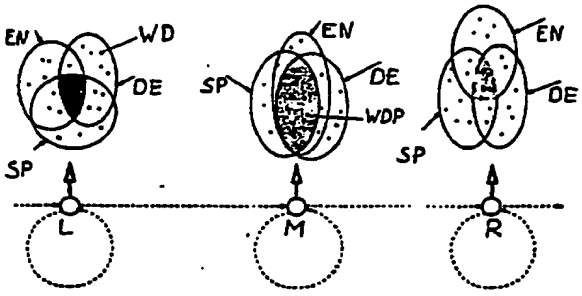
(72) Erfinder:
Köhler, Joachim, 80634 München, DE

(56) Für die Beurteilung der Patentfähigkeit in Betracht gezogene Druckschriften:

WO 95 02 879 A1
DIGALAKIS, V., SANKAR, A., BEAUFAYS, F.: »Training Data Clustering For Improved Speech Recognition.« In Proc. EUROSPEECH '95, pages 503-506, Madrid 1995;
DALSGAARD, F., ANDERSEN, O.: »Identification of Mono and Poly-phonemes using acousitc-phonetic Features derived by a self-organising Neural Network.« In Proc. ICSLP '92, pages 547-550, Banff 1992;
HAUENSTEIN, A., MARSCHALL, E.: »Methods for Improved Speech Recognition Over the Telephone Lines.« In Proc. ICASSP '95, pages 425-428, Detroit 1995;
HIERONYMUS, J.L.: »ASCII Phonetic Symbols for the World's Languages: Worldbet.« preprint, 1993;
LADEFOGED, P.: »A Course in Phonetics.« Harcourt Brace Jovanovich, San Diego 1993;
DALSGAARD, P., ANDERSEN, O., BARRY, W.: »Data- driven Identification of Poly- and Mono-phonemes for four European Languages.« In Proc. EUROSPEECH '93, pages 759-762, Berlin 1993;
COLE, A., MUTHUSAMY, Y.K., OSHIKA, B.T.: »The OGI Multilanguage Telephone Corpus.« In Proc. IC-SLP '92, pages 895-898, Banff 1992;

(54) Verfahren zur Mehrsprachenverwendung eines hidden Markov Lautmodelles in einem Spracherkennungssystem

(57) Mit der Erfindung wird eine Methode zur Bestimmung der Ähnlichkeiten von Lauten über verschiedene Sprachen hinweg angegeben. Weiterhin wird ein neuer Ansatz zur hidden Markov Modellierung von multilingualen Phonemen angegeben. Bei der vorgeschlagenen Methode zur akustisch phonetischen Modellierung werden sowohl sprachspezifische als auch sprachunabhängige Eigenschaften bei der Zusammenfassung der Wahrscheinlichkeitsdichten für unterschiedliche hidden Markov Lautmodelle in verschiedenen Sprachen angegeben.



DE 196 36 739 C1

DE 196 36 739 C1

Beschreibung

Die Erfindung bezieht sich auf hidden Markov Modelle für Spracherkennungssysteme, w bei ein solches Modell für mehrere Sprachen herangezogen werden soll, indem die akustischen und phonetischen Ähnlichkeiten zwischen den unterschiedlichen Sprachen ausgenutzt werden.

Ein Spracherkennungssystem für mehrere Sprachen ist aus der WO 95/02879 A1 bekannt.

Bei der Spracherkennung besteht ein großes Problem darin, daß für jede Sprache in welche die Spracherkennungstechnologie eingeführt werden soll, neue akustisch phonetische Modelle trainiert werden müssen um eine Länderanpassung durchführen zu können. Meistens werden bei gängigen Spracherkennungssystemen hidden Markov Modelle zur Modellierung der sprachspezifischen Laute verwendet. Aus diesen statistisch modellierten Lautmodellen werden im Anschluß akustische Wortmodelle zusammengefügt, welche während eines Suchprozesses beim Spracherkennungsvorgang erkannt werden. Zum Training dieser Lautmodelle werden sehr umfangreiche Sprachdatenbanken benötigt, deren Sammlung und Aufbereitung einen äußerst kosten- und zeitintensiven Prozeß darstellt. Hierdurch entstehen Nachteile bei der Portierung einer Spracherkennungstechnologie von einer Sprache in eine weitere Sprache, da die Erstellung einer neuen Sprachdatenbank einerseits eine Verteuerung des Produktes bedeutet und andererseits eine zeitliche Verzögerung bei der Markteinführung bedingt.

In gängigen erwerbbaren Spracherkennungssystemen werden ausschließlich sprachspezifische Modelle verwendet. Zur Portierung dieser Systeme in eine neue Sprache werden umfangreiche Sprachdatenbanken gesammelt und aufbereitet. Anschließend werden die Lautmodelle für die neue Sprache mit diesen gesammelten Sprachdaten von Grund auf neu trainiert.

Um den Aufwand und die Zeitverzögerung bei der Portierung von Spracherkennungssystemen in unterschiedliche Sprachen zu verringern, sollte also untersucht werden, ob einzelne Lautmodelle für die Verwendung in verschiedenen Sprachen geeignet sind. Hierzu gibt es in [2] bereits Ansätze mehrsprachige Lautmodelle zu erstellen und diese bei der Spracherkennung in den jeweiligen Sprachen einzusetzen. Dort werden auch die Begriffe Poly- und Monophoneme eingeführt. Wobei Polyphoneme Laute bedeuten, deren Lautbildungseigenschaften über mehrere Sprachen hinweg ähnlich genug sind, um gleichgesetzt zu werden. Mit Monophonemen werden Laute bezeichnet, welche sprachspezifische Eigenschaften aufweisen. Um für solche Entwicklungsarbeiten und Untersuchungen nicht jedesmal neue Sprachdatenbanken trainieren zu müssen, stehen solche schon als Standard zur Verfügung [6], [4], [7]. Ein weiterer Stand der Technik zur mehrsprachigen Verwendung von Lautmodellen ist nicht bekannt.

Die der Erfindung zugrundeliegende Aufgabe besteht demnach darin, ein Verfahren zur Mehrsprachenverwendung eines hidden Markov Lautmodelles in einem Spracherkennungssystem anzugeben, durch welches der Portierungsaufwand von Spracherkennungssystemen in eine andere Sprache minimiert wird, indem die Parameter in einem multilingualen Spracherkennungssystem reduziert werden.

Diese Aufgabe wird gemäß den Merkmalen der Patentansprüche 1 und 6 gelöst.

Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

Ein besonderer Vorteil des erfindungsgemäßen Verfahrens besteht darin, daß ein statistisches Ähnlichkeitsmaß angegeben wird, welches es erlaubt, aus einer gegebenen Anzahl von verschiedenen Lautmodellen für ähnliche Laute in unterschiedlichen Sprachen dasjenige Lautmodell auszuwählen, welches in seiner Charakteristik alle zur Verfügung stehenden Merkmalsvektoren der jeweiligen Laute am besten beschreibt.

Besonders vorteilhaft wird als Maß für die Auswahl des besten hidden Markov Modelles für unterschiedliche Lautmerkmalsvektoren der logarithmische Wahrscheinlichkeitsabstand zwischen den jeweiligen hidden Markov Modellen und einem jeden Merkmalsvektor ermittelt. Hierdurch wird ein Maß zur Verfügung gestellt, welches experimentelle Befunde bezüglich der Ähnlichkeit von einzelnen Lautmodellen und deren Erkennungsraten widerspiegelt.

Besonders vorteilhaft wird als Maß für die Beschreibung eines möglichst repräsentativen hidden Markov Lautmodelles nach der Erfindung der arithmetische Mittelwert der logarithmischen Wahrscheinlichkeitsabstände zwischen jedem hidden Markov Modell und den jeweiligen Merkmalsvektoren gebildet, da hierdurch ein symmetrischer Abstandswert erhalten wird.

Vorteilhaft wird das erfindungsgemäße Beschreibungsmaß für die repräsentative Eigenschaft eines hidden Markov Modells zur Beschreibung von Lauten in unterschiedlichen Sprachen dadurch gebildet, daß die erfindungsgemäßen Gleichungen 1 bis 3 angewendet werden, da hierdurch ein geringer Rechenaufwand entsteht.

Besonders vorteilhaft wird für die erfindungsgemäße Anwendung eines Beschreibungsmaßes eine Schrankenbedingung vorgegeben, mit der eine Erkennungsrate des repräsentierenden hidden Markov Modells eingestellt werden kann.

Besonders vorteilhaft wird durch das erfindungsgemäße Verfahren der Speicheraufwand für eine Sprachbibliothek reduziert, da ein Modell für mehrere Sprachen verwendet werden kann und ebenfalls der Portierungsaufwand von einer Sprache in die andere minimiert, was einen reduzierten Zeitaufwand für die Portierung bedingt. Ebenso vorteilhaft wird ein geringerer Rechenaufwand bei der Viterbi-Suche ermöglicht, da beispielsweise bei mehrsprachigen Eingabesystemen weniger Modelle überprüft werden müssen.

Besonders vorteilhaft werden bei der Erfindung besondere hidden Markov Modelle zur Verwendung in mehrsprachigen Spracherkennungssystemen generiert. Durch die erfindungsgemäße Vorgehensweise können hidden Markov Lautmodelle für Laute in mehreren Sprachen zu Polyphonem-Modellen zusammengefaßt werden. Hierzu werden Überlappungsbereiche der verwendeten Standardwahrscheinlichkeitsdichteverteilungen bei den unterschiedlichen Modellen untersucht. Zur Beschreibung des Polyphonem-Modelles kann eine beliebige Anzahl von identisch bei den unterschiedlichen Modellen verwendeten Standardwahrscheinlichkeitsdichteverteilungen herangezogen werden. Experimentelle Befunde haben gezeigt, daß vorteilhaft auch mehrere Standardverteilungen aus unterschiedlichen Sprachmodellen verwendet werden können, hne daß di hierdurch

bewirkte Verwischung der einzelnen Sprachcharakteristika zu einer signifikant niedrigeren Erkennungsrate beim Einsatz dieses Modells führen würde. Als besonders vorteilhaft hat sich hier der Abstandsschwellenwert fünf zwischen ähnlichen Standardwahrscheinlichkeitsverteilungsdichten bewährt.

Besonders vorteilhaft werden beim Einsatz des erfindungsgemäßen Verfahrens die hidden Markov Modelle mit drei Zuständen aus Anlaut, Mittellaut und Ablaut modelliert, da hierdurch eine hinreichende Genauigkeit bei der Beschreibung der Laute erzielt wird und der Rechenaufwand bei der Erkennung in einem Spracherkenner gering bleibt.

Fig. 1 zeigt dabei beispielhaft den Aufbau eines einzigen Multilingualen Phonemes. In diesem Fall ist es das Phonem M was dargestellt wird. Die Zahl der Wahrscheinlichkeitsdichten und die Erkennungsrate für dieses Phonem sind in Tabelle 4 angegeben.

Thr.	#densit(a,b,c).	Engl.[%]	Germ.[%]	Span.[%]
0	341(0 0 341)	46.7	44.7	59.4
2	334(0 14 327)	45.0	46.4	57.5
3	303(27 34 280)	48.0	45.8	57.5
4	227(106 57 187)	50.9	44.1	58.7
5	116(221, 48,72)	49.3	43.1	57.0
6	61(285, 22, 34)	41.2	38.6	50.4

In Fig. 1 ist der Anlaut L, der Mittel laut M und der Ablaut R des Phonem-Modelles dargestellt. Für die unterschiedlichen Sprachen Englisch EN, Deutsch DE und Spanisch SP sind die Schwerpunkte der Wahrscheinlichkeitsdichteverteilungen der einzelnen verwendeten Standardwahrscheinlichkeitsdichten eingetragen und als WD gekennzeichnet. Hier ist beispielsweise ein hidden Markov Modell aus drei Teilzuständen dargestellt. Die Erfindung soll jedoch nicht lediglich auf solche hidden Markov Modelle beschränkt werden, obwohl diese unter Berücksichtigung des Kriteriums, das ein minimaler Rechenaufwand der Erkennung durchgeführt werden soll ein gewisses Optimum darstellen. Die Erfindung kann ebenso auf hidden Markov Modelle angewendet werden, die eine andere Anzahl von Zuständen aufweisen. Durch die Erfindung soll insbesondere erreicht werden, daß der Portierungsaufwand bei der Portierung von Spracherkennungssystemen in eine andere Sprache reduziert wird und daß die verwendeten Rechenressourcen durch Reduktion der zugrundeliegenden Parameter möglichst gering gehalten werden. Beispielsweise können durch derartige Spracherkennungssysteme begrenzte Hardwareerfordernisse besser erfüllt werden, insbesondere wenn ein- und dasselbe Spracherkennungssystem für Mehrsprachenanwendung in einem Gerät zur Verfügung gestellt werden soll.

Zunächst sollte um das Ziel der Erfindung zu erreichen, die Ähnlichkeiten von Lauten in unterschiedlichen Sprachen auszuschöpfen und beim Modellieren zu berücksichtigen, beachtet werden, daß sich die Phoneme in verschiedenen Sprachen unterscheiden können. Die Gründe hierfür bestehen vor allen Dingen in:

- Unterschiedlichen phonetischen Kontexten, wegen der unterschiedlichen Phonemsätze in den verschiedenen Sprachen;
- unterschiedlichen Sprechweisen;
- verschiedenen prosodischen Merkmalen;
- unterschiedlichen allophonischen Variationen.

Ein besonders wichtiger Aspekt, welcher dabei zu berücksichtigen ist, besteht im Prinzip der genügenden wahrnehmungstechnischen Unterscheidbarkeit der Phoneme [5]. Dies bedeutet, daß einzelne Laute in verschiedenen Sprachen akustisch unterscheidbar gehalten werden, so daß es für den einzelnen Zuhörer leichter ist sie voneinander zu separieren. Da aber jede einzelne Sprache einen unterschiedlichen Phonemschatz hat, werden die Grenzen zwischen zwei ähnlichen Phonemen in jeder einzelnen Sprache sprachspezifisch festgelegt. Aus diesen Gründen hat die Ausprägung eines bestimmten Lautes eine sprachspezifische Komponente.

Bevorzugt werden die Phoneme mittels kontinuierlichen dichten hidden Markov Modellen (CD-HMM) modelliert [3]. Als dichte Funktionen werden häufig Laplace-Mischungen benutzt. Bevorzugt besteht dabei jedes einzelne Phonem aus drei Zuständen von links nach rechts gerichteten HMM. Die akustischen Merkmalsvektoren bestehen dabei beispielsweise aus 24 mel-skalierten cepstral, 12 delta cepstral, 12 delta delta cepstral, Energie, delta-Energie und delta delta-Energie-Koeffizienten. Beispielsweise wird als Länge des Untersuchungszeitfensters 25 ms gewählt, wobei die Rahmenabstände 10 ms zwischen den einzelnen Rahmen betragen. Aus Gründen der begrenzten Größe des Sprachkorpus werden bevorzugt lediglich kontextunabhängige Phoneme generiert. Als bes nders repräsentatives Phoneminventar wurde jenes aus [4] gewählt.

Die Idee der Erfindung besteht dabei darin, daß zum einen ein Ähnlichkeitsmaß zur Verfügung gestellt wird, um aus standardmäßig verfügbaren Sprachphonembibliotheken für unterschiedliche Sprachen jenes hidden Markov Modell auswählen zu können, welches den Merkmalsvektoren, die aus den unterschiedlichen Lautmodellen der unterschiedlichen Sprachen abgeleitet werden, am nächsten kommt. Hierdurch ist es möglich, die Ähnlichkeiten zweier Phonem-Modelle zu ermitteln und über dieses Ähnlichkeitenmaß basierend auf der

Differenz der Log-Likelihood-Werte zwischen den Lautrealisierungen und Lautmodellen eine Aussage zu treffen, ob es sich lohnt, einen Laut für mehrere Sprachen gemeinsam zu modellieren, bzw. ein betreffendes schon bestehendes hidden Markov Modell für die Modellierung des Lautes in mehreren Sprachen zu verwenden. Hierdurch wird die Zahl der Parameter, welche bei der Spracherkennung zu berücksichtigen sind reduziert, indem die Zahl der zu untersuchenden hidden Markov Modelle reduziert wird.

Ein zweiter Lösungsansatz der Erfindung besteht darin, ein spezielles Polyphonem-Modell zur Modellierung eines Lautes in mehreren Sprachen zu erstellen. Hierzu werden zunächst beispielsweise drei Lautsegmente, in Form eines Anlautes, Mittellautes und Ablautes gebildet, deren Zustände aus mehreren Wahrscheinlichkeitsdichtefunktionen den sogenannten Mischverteilungsdichten mit den dazugehörigen Dichten bestehen. Diese Dichten der über verschiedenen Sprachen ähnlichen Lautsegmente werden zu einem multilingualen Codebuch zusammengefaßt. Somit teilen sich Lautsegmente verschiedener Sprachen die gleichen Dichten. Während das Codebuch für mehrere Sprachen gleichzeitig benutzt werden kann, werden beispielsweise die Gewichte, mit denen die Dichten gewichtet werden für jede Sprache getrennt ermittelt.

Zur Bildung eines geeigneten Ähnlichkeitsmaßes werden bevorzugt hidden Markov Modelle mit drei Zuständen herangezogen. Das Abstands- oder Ähnlichkeitsmaß kann dabei benutzt werden um mehrere Phonem-Modelle zu einem multilingualen Phonem-Modell zusammenzufassen oder diese auf geeignete Weise zu ersetzen. Hierdurch kann ein multilingualer Phonemschatz entwickelt werden. Bevorzugt wird zur Messung des Abstandes bzw. zur Bestimmung der Ähnlichkeit von zwei Phonem-Modellen des selben Lautes aus unterschiedlichen Sprachen eine Meßgröße verwendet, welche auf der relativen Entropie basiert [1]. Während des Trainings werden dabei die Parameter der gemischten Laplacedichteverteilungen der Phonem-Modelle bestimmt. Weiterhin wird für jedes Phonem ein Satz von Phonemtokens X als Merkmalsvektor aus einem Test- oder Entwicklungssprachkorpus extrahiert. Diese Phoneme können dabei durch ihr international genormtes phonetisches Etikett markiert sein. Gemäß der Erfindung werden zwei Phonem-Modelle λ_i und λ_j und ihre zugehörigen Phonemtoken X_i und X_j zur Bestimmung des Ähnlichkeitsmaßes zwischen diesen unterschiedlichen Phonemen wie folgt behandelt.

$$d(\lambda_i,\lambda_j) = \log p(X_i|\lambda_i) - \log p(X_i|\lambda_j) \quad (1)$$

Dieses Abstandsmaß kann als Log-Likelihood-Abstand angesehen werden, welcher darstellt wie gut zwei verschiedene Modelle zu dem selben Merkmalsvektor X_I passen. Demgemäß wird der Abstand zwischen den beiden Modellen λ_i und λ_j gemäß:

$$d(\lambda_j,\lambda_i) = \log p(X_j|\lambda_j) - \log p(X_j|\lambda_i) \quad (2)$$

bestimmt. Um einen symmetrischen Abstand zwischen diesen beiden Phonem-Modellen zu erhalten, wird dieser bevorzugt gemäß

$$d(\lambda_j;\lambda_i) = \frac{1}{2}\left(d(\lambda_i,\lambda_j) + d(\lambda_j;\lambda_i)\right) \quad (3)$$

bestimmt. Anhand von experimentellen Befunden konnte festgestellt werden, daß sich durchaus einige Phonem-Modelle aus anderen Sprachen besser für die Verwendung in einem deutschen Spracherkennungssystem eignen, als ein deutsches Phonem-Modell. Beispielsweise gilt dies für die Phoneme k, p und N. Für diese Phoneme eignet sich das englische Phonem-Modell besser als das deutsche. Während beispielsweise ein großer Unterschied zwischen dem deutschen und dem englischen Modell über den Umlaut aU beobachtet wurde, was bedeutet, daß für beide Laute ein unterschiedliches Symbol im multilingualen Phonemschatz eingeführt werden sollte. Andererseits konnte für den Umlaut aI im deutschen und im englischen eine große Ähnlichkeit festgestellt werden, das bedeutet, daß lediglich ein Phonem-Modell für beide Sprachen gleich gut Verwendung finden kann. Ausgehend davon sollte für jedes Symbol eines multilingualen Phonemschatzes ein separates statistisches Modell erzeugt werden. In [6] wurden Polyphoneme als solche Phoneme bezeichnet, die ähnlich genug sind, um in verschiedenen Sprachen als ein einziges Phonem modelliert zu werden. Ein Nachteil dieser Vorgehensweise besteht darin, daß für die sprachspezifische Erkennung der vollständige akustische Raum des Polyphonems verwendet wird. Die Erfindung hat es jedoch zum Ziel, die sprachabhängigen und die sprachspezifischen akustischen Eigenschaften eines multilingualen Modells zu kombinieren. Gemäß der Erfindung sollen in einem Polyphonem-Modell solche Bereiche des akustischen Raumes eingegrenzt werden, in denen sich die verwendeten Wahrscheinlichkeitsdichten der einzelnen Phoneme überlappen. Hierzu wird z. B. eine gruppierende Verdichtungstechnik (agglommerative density clustering technique) eingesetzt, um gleiche oder ähnliche Ausprägungen eines Phonems zu reduzieren. Besonders wichtig ist es dabei zu beachten, daß lediglich die Dichten der korrespondierenden Zustände der einzelnen hidden Markov Modelle in den Phonemen zusammengefaßt werden dürfen.

In Fig. 1 ist dabei zu erkennen, daß die jeweiligen Dichten für die einzelnen Zustände L, M und R in den eingegrenzten Regionen enthalten sind. Während identische Dichten über die einzelnen Sprachen EN, DE, und SP verteilt sind, variieren die Mischungsgewichte sprachabhängig. Bei dieser Bewertung sollte jedoch auch berücksichtigt werden, daß spezifische Ausprägungen eines Phonems in verschiedenen Sprachen in unterschiedlicher Häufigkeit auftreten.

Die Zusammenfassung der unterschiedlichen Wahrscheinlichkeitsdichten kann dabei mit einem unterschiedlichen Abstandsschwellenwert für die Wahrscheinlichkeitsdichten bei der Dichtehäufung (density clustering)

durchgeführt werden. Beispielsweise wurde mit einem Abstandsschwellenwert von fünf die Zahl der verwendeten Dichten um einen Faktor 3 gegenüber dem Ausgangszustand reduziert, ohne damit eine entscheidende Verschlechterung bei der Spracherkennungsrate einher ging. In diesem Fall wurden 221, 48 und 72 von den ursprünglichen 341 Ausgangsdichten für jeweils die Polyphonem-Region, die Zweisprachen-Region und die Einsprachen-Region zusammengefaßt. In Fig. 1 ist eine solche Polyphonemregion als Schnittmenge der Kreise für die einzelnen Sprachen dargestellt. Beim Mittellaut M des dargestellten hidden Markov Modells ist beispielsweise eine Wahrscheinlichkeitsdichte in einer solchen Region als WDP bezeichnet. Die Erkennungsraten für ein komplettes multilinguales Spracherkennungssystem sind dabei in Spalte 4 und 5 der Tabelle 2 als ML1 und ML2 angegeben.

Language	#Tokens	LDP[%]	ML1[%]	ML2[%]
English	21191	39.0	37.3	37.0
German	9430	40.0	34.7	37.7
Spanish	9525	53.9	46.0	51.6
Total	40146	42.8	38.8	40.8

Während bei der ersten Untersuchung ML1 die konventionelle Polyphonem-Definition aus [6] verwendet wurde, was bedeutet, daß der komplette akustische Bereich des Polyphonem-Modells bestehend aus der äußeren Kontur der Sprachbereiche in Fig. 1, für die Erkennung verwendet wurde, benutzt die erfindungsgemäße Methode lediglich einen Teilbereich daraus. Indem die teilweise Überlappung der einzelnen Sprachbereiche für die einzelne Modellierung des Polyphonem-Modells herangezogen wird, ist beispielsweise eine Verbesserung von 2% erzielbar, wie dies in Tabelle 2 in der Spalte für ML2 dargestellt ist.

Literatur

[1] V. Digalakis A. Sankar, F. Beaufays.: "Training Data Clustering For Improved Speech Recognition.", In Proc. EUROSPEECH '95, pages 503—506, Madrid, 1995;
[2] P. Dalsgaard and O. Andersen.: "Identification of Mono- and Poly-phonemes using acoustic-phonetic Features derived by a self-organising Neural Network.", In Proc. ICSLP '92, pages 547—550, Banff, 1992;
[3] A. Hauenstein and E. Marschall.: "Methods for Improved Speech Recognition Over the Telephone Lines.", In Proc. ICASSP '95, pages 425—428, Detroit, 1995;
[4] J. L. Hieronymus.: "ASCII Phonetic Symbols for the World's Languages: Worldbet.", preprint, 1993;
[5] P. Ladefoged: A Course in Phonetics.", Harcourt Brace Jovanovich, San Diego, 1993;
[6] P. Dalsgaard O. Andersen and W. Barry.: "Data-driven Identification of Poly- and Mono-phonemes for four European Languages.", In Proc. EUROSPEECH '93, pages 759—762, Berlin, 1993;
[7] A. Cole Y.K. Muthusamy and B.T. Oshika.: "The OGI Multilanguage Telephone Speech Corpus.", In Proc. IC-SLP '92, pages 895—898, Banff, 1992.

Patentansprüche

1. Verfahren zur Mehrsprachen Verwendung eines hidden Markov Lautmodelles in einem Spracherkennungssystem,
   a) bei dem ausgehend von mindestens einem ersten Merkmalsvektor für einen ersten Laut (L,M,R) in einer ersten Sprache (SP,EN,DE) und von mindestens einem zweiten Merkmalsvektor für einen vergleichbar gesprochenen zweiten Laut in mindestens einer zweiten Sprache (DE,SP,EN) und deren zugehörigen ersten und zweiten hidden Markov Lautmodellen, ermittelt wird welches der beiden hidden Markov Lautmodelle (L,M,R) beide Merkmalsvektoren besser beschreibt,
   b) und bei dem dieses hidden Markov Lautmodell (L,M,R) für die Modellierung des Lautes in mindestens beiden Sprachen (SP,EN,DE) verwendet wird.
2. Verfahren nach Anspruch 1, bei dem als Maß für die Beschreibung eines Merkmalsvektors durch ein hidden Markov Lautmodell (L,M,R) der logarithmische Wahrscheinlichkeitsabstand als log likelihood distance zwischen jedem hidden Markov Lautmodell und mindestens einem Merkmalsvektor gebildet wird, wobei eine kürzerer Abstand eine bessere Beschreibung bedeutet.
3. Verfahren nach Anspruch 2, bei dem als Maß für die Beschreibung der Merkmalsvektoren durch die hidden Markov Lautmodelle der arithmetische Mittelwert der logarithmischen Wahrscheinlichkeitsabstände bzw. der log likelihood distances zwischen jedem hidden Markov Lautmodell (L,M,R) und jedem jeweiligen Merkmalsvektor gebildet wird, wobei eine kürzerer Abstand eine bessere Beschreibung bedeutet.
4. Verfahren nach Anspruch 3, bei dem das erste hidden Markov Lautmodell (L,M,R) von einem Phonem λ_i und das zweite hidden Markov Lautmodell von einem Phonem λ_j verwendet wird und bei dem als erste und zweite Merkmalsvektoren X_i und X_j verwendet werden, wobei der logarithmische Wahrscheinlichkeitsabstand zum ersten Merkmalsvektor gemäß

$d(\lambda_i,\lambda_j) = l\ g\ p(X_i|\lambda_i) - \log p(X_i|\lambda_j)$ (1)

bestimmt wird und der logarithmische Wahrscheinlichkeitsabstand zum zweiten Merkmalsvektor gemäß

$d(\lambda_j,\lambda_j) = \log p(X_j|\lambda_j) - \log p(X_j|\lambda_i)$ (2)

bestimmt wird, wobei zur Erzielung eines symmetrischen Abstandsmaßes der arithmetische Mittelwert zu

$$d(\lambda_j;\lambda_i) = \frac{1}{2}\left(d(\lambda_i,\lambda_j) + d(\lambda_j;\lambda_i)\right) \quad (3)$$

5. Verfahren nach Anspruch 4, bei dem dieses hidden Markov Lautmodell (L,M,R) für die Modellierung des Lautes in mindestens beiden Sprachen nur verwendet wird, falls $d(\lambda_j;\lambda_i)$ eine festgelegte Schrankenbedingung erfüllt.

6. Verfahren zur Mehrsprachenverwendung eines hidden Markov Lautmodelles in einem Spracherkennungssystem,

a) bei dem ausgehend von mindestens einem ersten hidden Markov Lautmodell (L,M,R) für einen ersten Laut in einer ersten Sprache (SP,EN,DE) und von mindestens einem zweiten hidden Markov Lautmodell (L,M,R) für einen vergleichbar gesprochenen zweiten Laut in mindestens einer zweiten Sprache (DE,SP,EN), ein Poly Phonem Modell derart gebildet wird, daß die für die Modellierung des ersten und zweiten hidden Markov Lautmodelles (L,M,R) verwendeten Standardwahrscheinlichkeitsverteilungen (WD) bis zu einem festgelegten Abstandsschwellenwert, der angibt bis zu welchem maximalen Abstand zwischen zwei Standardwahrscheinlichkeitsverteilungen (WD) diese zusammengefügt werden sollen zu jeweils einer neuen Standardwahrscheinlichkeitsverteilung (WDP) zusammengefügt werden und lediglich die zusammengefügten Standardwahrscheinlichkeitsverteilungen das Poly Phonem Modell charakterisieren

b) und bei dem dieses Poly Phonem Modell für die Modellierung des Lautes in mindestens beiden Sprachen (DE,SP,EN) (L,M,R) verwendet wird.

7. Verfahren nach Anspruch 6, bei dem als Abstandsschwellenwert 5 festgelegt wird.

8. Verfahren nach einem der vorangehenden Ansprüche bei dem hidden Markov Lautmodelle mit drei Zuständen verwendet werden, welche aus den Lautsegmenten Anlaut, Mittellaut und Ablaut gebildet werden.

Hierzu 1 Seite(n) Zeichnungen

– Leerseite –

Nummer: DE 196 36 739 C1
Int. Cl.[6]: G 10 L 5/06
Veröffentlichungstag: 3. Juli 1997

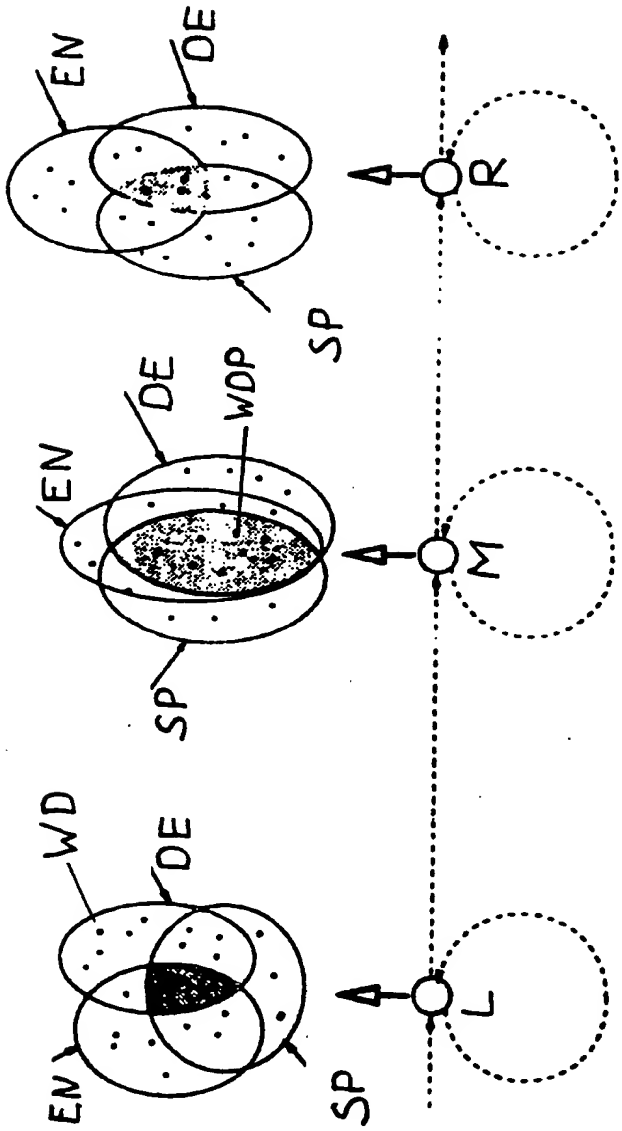


FIG 1